

# 표본중도절단에 따른 통계학적 문제와 교정방법에 관한 고찰

— 소비자분야 연구에의 적용을 중심으로 —

Statistical Problems Caused by Sample Censoring and Their Solutions

— Focused on the application to consumer research —

숙명여자대학교 소비자경제학과

강사 나 명균

Dept. of Consumer Economics, Sook Myung Women's University

Lecturer : Myungkyun Nah

## 〈목 차〉

- I. 서론
- II. 표본중도절단에 따르는 통계학적 문제들
- III. 중도절단문제에 대한 통계학적 해결안
- IV. 표본중도절단 교정의 예
- V. 표본중도절단을 교정한 기타 연구들
- VI. Heckman의 교정방법에서 고려할 문제
- VII. 결론

## 〈Abstract〉

This paper discusses the bias that results from using nonrandomly selected samples of consumer research. A two stage system(maximum likelihood probit analysis and ordinary least square analysis) is a solution to sample selection bias. Empirical results show that correcting for sample selection bias improves the validity of consumer research results.

## I. 서 론

학문을 연구하는데 있어 통계학적 문제는 종종 전 공분야의 영역을 뛰어 넘는다. 그러나 아직까지는 학 문간의 연결이 잘 이루어지고 있지 않기 때문에 통 계학적 문제의 해결책을 발견한 경우는 그 분야내에 서의 업적으로만 남는 경우가 종종 있으며, 같은 원 리의 통계처리도 학문간의 용어의 차이로 상호교환 을 어렵게 하고있다. 표본중도절단(sample censoring) 의 경우도 이러한 학문적 편협주의의 한 예로 볼 수 있다.

표본중도절단은 표본이 모집단을 충실히 대표해야 하는 확률표본의 원리에 어긋난다. 이것은 연구표본 의 일부에 대해서 독립변수의 알려진 값에 상응하는 종속변수에 대한 값을 관찰할 수 없을때 일어나는데 일반적으로 직접중도절단(direct censoring)과 간접중 도절단(indirect censoring)으로 나누고 있다(Berk, 1983). 직접중도절단은 종속변수의 값이 어느 수준 이하이기 때문에 표본에서 제외시킬때 일어난다. 예 를 들어 가전제품의 지출비용에 영향을 주는 변수들 에 관하여 연구조사하고자 할 경우, 조사기간 동안 에 가전제품을 구매하지 않은 가정은 직접중도절단 에 의해 표본에서 제외가 된다. 이 문제는 Tobit 분 석을 이용하여 교정할 수 있다.

간접중도절단은 표본선택이 연구자의 관심대상에 어떤 역할을 하지않는 외생변수(exogenous variable) 의 값에 근거하여 이루어질때 일어난다. 이에 관한 대표적인 예는 기혼여성의 임금추정이다. 취업주부 의 경우에는 시장임금을 관찰할 수가 있으나, 전업 주부라면 임금관찰이 불가능하다. 이러한 표본을 연 구분석대상에서 삭제하는 것이 표본중도절단으로 여 성의 취업을 결정하는 요소들이 대부분의 경우 그녀 가 받는 시장임금에 영향을 주는 요소들과 같기 때 문이다. 이외에 재택근무자(home-based worker)의 임금추정이나 소비자 불평행동을 연구하기 위해 소 비자불만창구를 이용했던 사람들을 대상으로 조사대 상자를 선별할 경우(Kolodinsky, 1992) 등도 같은 문 제를 갖고 있다. 즉 실증적 연구에서 조사대상자들 이 나타내는 의미를 통계방법상의 오류로 그릇되게

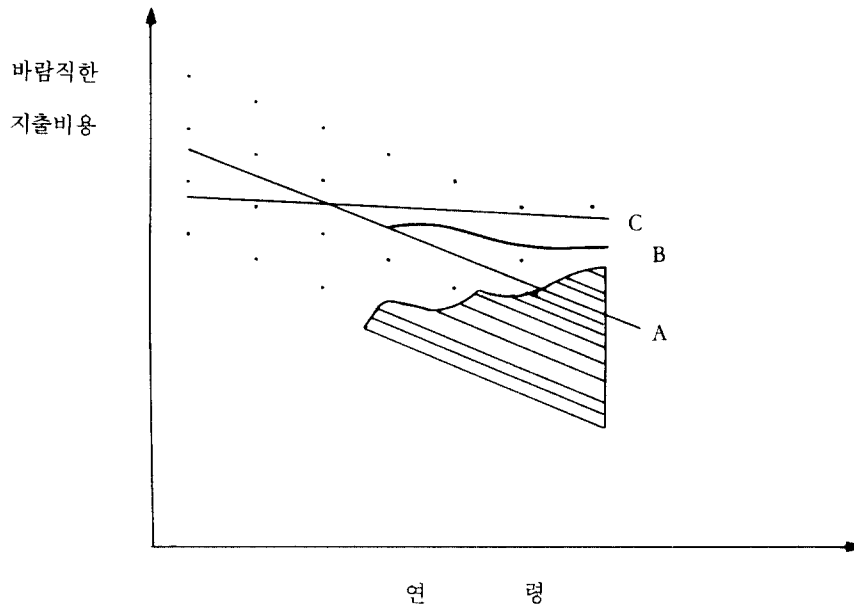
나타낸다는 점이다. 이러한 문제에 대해서는 경제학 자들을 중심으로 많은 관심을 보여왔으나 다른 분야 에서는 소홀히 하는 경향이 있다. 최근들어 가정관 리 분야나 소비자 분야의 실증연구에 외국문헌을 중 심으로 표본선택편의의 교정을 위해 Probit 분석을 이용한 연구들이 나타나고 있다.

본 연구에서는 실증연구에 있어 자료처리 방법상 의 문제점을 다시한번 고려해 보고 자료가 나타내는 의미를 정확히 파악할 수 있는 방법을 제시하며 연 구결과를 가능하면 일반화시킬 수 있도록 하는데 도 움이 되고자 한다. 따라서 본 연구에서는 간접중도 절단에 의한 통계학적 문제와 그 해결방안에 관하여 살펴보고 교정된 실증연구결과와 교정전의 연구결과 를 비교하며 실제로 어떠한 연구형태에 응용될 수 있는지를 제시하는데 그 목적이 있다.

## II. 표본중도절단에 따르는 통계학적 문 제들

실증연구자료를 회귀분석할 때 표본중도절단을 한 다면 몇가지 통계학적 문제가 일어난다(Zick, 1985). 첫째, 절단된 자료로 추정된 회귀계수는 모집단에 대 한 회귀선의 원래의 기울기(true slope)를 그릇되게 나타낸다. 많은 학자들이 이러한 문제점을 인식하기 때문에 제한된 표본으로 이루어진 실증연구의 외적 타당성(external validity)에 관하여 부제나 제한점에 미리 표시하는 것을 상당히 자주 보게된다. 그러나 그 연구결과가 다른연구에 인용될때 제한점이나 부 세를 명시한채 결과만을 인용하는 사례가 많아 원래 의 결과에서 벗어난 의미로 전달되는 결과를 낳게된 다. 둘째, 추정된 회귀계수가 대체로 편의될(biased) 가능성이 있다. 이 문제점은 <그림 1>에서 잘 설명될 수 있다.

예를 들어 최근에 내구재를 구입한 조사대상자들 의 바람직한 내구재 지출비용에 대한 자료를 분석한 결과, 바람직한 지출비용수준은 응답자의 연령과 관 계가 있어서 소비자의 연령이 높아질수록 내구재에 대한 지출이 줄어든다고 나타나 있다. 또한 연령이 높을수록 최근에 내구재의 구입경험이 적은 것으로

〈그림 1〉 내구재지출비용과 연령과의 관계<sup>a</sup>

나타났다. 이러한 분석결과는 최근에 내구재 구입이 없었던 고연령 소비자들이 표본에서 절단되었기 때문으로 절단부분은 그림에서 사선으로 나타나 있다. 모든 자료가 이용가능하다면 바람직한 내구재 지출비용과 연령사이의 적합한 회귀선은 A일 것이다. 그러나 일부자료의 조직적인 절단으로 인하여 바람직한 지출비용과 연령사이의 관계는 비선형적 곡선인 B로 나타나 있다. 이러한 비선형적 관계를 직선형에 맞게 하려는 어떠한 노력도 조직적인 표기오차(specification error)를 범할 것이다. 즉 절단된 자료로 회귀분석을 하여 C선으로 추정되었다면, 저연령층에 대한 오차는 양수이고 고연령층에 대한 오차는 음수로 나타나서 조사대상자의 연령은 오차와 음의 상관관계가 있으므로 실제로는 확률표집오차(random sampling error)의 영향을 응답자의 연령 탓으로 잘못 인식하게 된다. 따라서 편의된 연구결과를 일

반화시킬 수 없을뿐만 아니라 최근 내구재를 구매한 소비자의 제한된 부분표본(subsample)의 전형으로조차도 이용할 수 없다. 다시말하면 간접중도절단은 실증연구의 외적타당성 뿐만 아니라 내적타당성(internal validity)까지도 위협하게 된다. 특히 다원적 모델에 의한 연구일 경우에는 표본중도절단에 의한 편의의 범위를 알아내기가 더욱 어렵기 때문에 주의가 필요하다."

또한 표본중도절단은 자료분석만이 아니라 질문지를 만들때 방법상의 부주의로도 일어난다(Berk, 1983). 예를들어 소비자 불평행동의 변화에 관심이 있는 연구자가 상품에 대한 불평을 나타냈던 사람들로만 조사대상자를 선별할 경우(Kolodinsky, 1992) 응답자에게 적용되지 않는 질문지를 읽을 필요가 없으므로 질문에 답하기가 훨씬 쉽다는 점에서는 바람직하나 동시에 이러한 선별은 간접중도절단의 문제

<sup>a</sup> Berk(1983)에서 인용

를 낼 수 있다. 결국 표본중도절단에 의한 문제는 실증연구에 있어서 중요한 자료분석상의 오류로 조사대상자들의 실태를 제대로 파악하지 못함으로써 잘못된 연구결과를 가져오게 한다.

### III. 중도절단문제에 대한 통계학적 해결안

Heckman(1979) 방법은 회귀식내에 표본중도절단에 의해 나타난 편이가 포함되었을때 이를 교정하는 도구변수(instrumental variable)를 이용하는 것이다. 이것을 남녀의 임금에 관한 연구를 예로 들어 살펴보면 다음과 같다.

남녀임금에 관한 연구는 취업한 남녀에 대한 임금 관찰만이 가능함을 나타내기 때문에 표본선택 문제를 두 방정식으로부터 생각해야 한다.

$$Y_{1i} = X_{1i}B_1 + U_{1i} \quad i=1, \dots, N \quad (1)$$

$$Y_{2i} = X_{2i}B_2 + U_{2i} \quad i=1, \dots, N \quad (2)$$

$Y_{1i}$  = 관심을 갖고있는 내생변수

(예, 남녀의 임금률)

$X_{1i}$  =  $Y_{1i}$ 에 영향을 준다고 가정된 외생변수들의 벡터

(예, 학력, 연령, 가족수 등)

$Y_{2i}$  =  $Y_{1i}$ 의 관찰을 위해 선택기준으로서 작용하는 내생변수 (예, 취업하고 있는 남녀)

$X_{2i}$  = 선택기준에 영향을 준다고 가정된 외생변수들의 벡터, 이 변수들의 일부는 또한  $X_{1i}$  벡터에도 포함된다 (예, 취업가능성에 영향을 주는 학력 등)

$N$  = 총표본수

등식 (1), (2)의 오차들은 회귀분석의 원리에 의해 각 등식에서의 오차의 기대값은 0 이지만 등식사이의 상관에 있는 공분산(covariance :  $\sigma_{12}$ )은 등식 (1)에서의 종속변수(남녀의 임금률)가 등식 (2)에 있는 종속변수(취업여부)에 의해 영향을 받는다는 것을 가리키고 있다. 따라서 중도절단된 표본에 대한 (1)의 등식은 다음과 같이 나타낼 수 있다.

$$E(Y_{1i} | X_{1i}, \text{선택기준}) = X_{1i}B_1 + E(U_{1i} | \text{선택기준}), \quad i=1, \dots, C \quad (3)$$

$C = N$  - 표본중도절단으로 잃어버린 표본수.

등식 (2)의 종속변수가 0보다 클 경우에만 등식 (1)에 있는 종속변수가 관찰될 수 있도록 중도절단 규칙이 규정하고 있으므로 (즉, 취업을 하고 있어야만 임금을 산출할 수 있으므로), 등식 (3)은 다음과 같이 쓸 수 있다.

$$E(Y_{1i} | X_{1i}, Y_{2i} > 0) = X_{1i}B_1 + E(U_{1i} | U_{2i} > -X_{2i}B_2) \quad (4)$$

$$E(U_{1i} | U_{2i} > -X_{2i}B_2) = \left( \frac{-\sigma_{12}}{(\sigma_2^2)^{1/2}} \right) \lambda_i \quad (5)$$

그리고,

$$\lambda_i = \frac{f\left(\frac{-X_{2i}B_2}{(\sigma_2^2)^{1/2}}\right)}{(1-F\left(\frac{-X_{2i}B_2}{(\sigma_2^2)^{1/2}}\right))} \quad (6)$$

등식 (6)에서  $f(\bullet)$ 는 정규분포로 가정된  $-X_{2i}B_2 / (\sigma_2^2)^{1/2}$ 의 확률밀도함수(probability density function)를 나타낸다.  $F(\bullet)$ 는 또한 정규분포로 가정된  $-X_{2i}B_2 / (\sigma_2^2)^{1/2}$ 의 누적밀도함수(cumulative density function)를 나타낸다. 이러한 결과로 등식 (4)는 다음과 같이 쓸 수 있다.

$$E(Y_{1i} | X_{1i} ; Y_{2i} > 0) = X_{1i}B_1 + \left( \frac{\sigma_{12}}{(\sigma_2^2)^{1/2}} \right) \lambda_i \quad (7)$$

위에서 오른쪽 마지막항은 자체선택편의(self selection bias)에 의한 것임을 나타낸다.  $\lambda$ 는 밀의 비율(Mill's ratio)의 역이거나 혹은 위험률(hazard rate)로 더욱 많이 알려져 있다. 그것은 위험부담을 안고 어떤 조건에 의해 본래의 중도절단되지않은 표본으로부터 일부표본을 제외시킬 가능성을 나타낸다(Tuma, 1982). 이렇게 볼때 Heckman은 중도절단에

따르는 문제를 누락된 독립변수에 대한 교정방법으로 바꾸어놓은 것이다. 다시말하면 등식 (7)에서  $\sigma_{12}$ 가 0 이거나(즉 등식 (1)과 (2)의 오차항이 상호 독립적인 경우)  $\lambda$ 가 0 이면(표본선택기준이 모집단을 충실히 대표하는 경우) 확률표본의 조건을 만족시키므로 선택편의를 교정할 필요가 없어진다.

#### IV. 표본중도절단 교정의 예

남녀의 임금을 추정하는 예에서 어떻게 표본중도절단이 일어나는지 그리고 그 결과에서는 어떠한 차이가 있는지 고려해보면 다음과 같다. 앞에서 설명한 바와 같이 취업하지 않은 남녀에 대해서는 임금을 산출할 수가 없기 때문에 표본중도절단을 하고 남아 있는 부분표본에 대해서 회귀모형을 추정하기가 쉽다. 이러한 방법으로 나타나는 편의의 정도를 알기위해, 표본중도절단을 교정한 경우와 하지 않은 경우의 추정결과를 비교해 보았다.

예를 든 연구(나명균, 1991)는 미국의 전국적인 표본조사 자료인 NLSY(National Longitudinal Survey

of Youth)를 이용하여 지속적인 전일제 근로자인 미국 남녀간의 임금격차에 대하여 Becker(1985)의 가설을 검증하였다. 연구의 종속변수는 조사대상자의 시간당 임금에 자연로그를 취한 것이고 독립변수는 인적자원변수들(교육, 직업경험, 직업훈련)과 시간외 근무, 연령, 현재의 직장이 단체교섭권을 가지고 있는지의 여부와 주거지역(도시와 시골, 남부와 기타 지역), 직장이 정부기관이나 혹은 사설기관인가의 여부, 6세이하 자녀의 유무, 부양가족수 등이 포함 되었다.

우선 표본중도절단에 대한 교정을 하지 않을 경우 회귀등식(등식 6)은 다음과 같을 것이다.

$$\begin{aligned} \text{LN(WAGE)} = & \text{HSGRAD} + \text{VOCED} + \text{COLGRAD} \\ & + \text{PROF} + \text{WEEKWORK} + \text{JOBTRN} \\ & + \text{OVERTIME} + \text{AGE} + \text{COLBAR} \\ & + \text{URBAN} + \text{SOUTH} + \text{PUBLIC} \\ & + \text{CHILD6} + \text{DEPENDNT} \quad (8) \end{aligned}$$

그러나 선택편의에 대한 교정은 몇가지 단계로 이

〈표 1〉 기혼 남녀가 지속적인 전일제 근로자일 가능성에 대한 Probit 분석

독립변수	기혼남성		기혼여성	
	B	t	B	t
HSGRAD	0.48***	3.94	0.45*	2.26
VOCED	0.45**	2.92	0.50*	2.32
COLGRAD	0.57***	3.57	0.25	1.11
PROF	-0.07	-0.34	0.44	1.50
CHILD6	0.28**	2.73	-0.16	-1.34
DEPENDNT	0.01	0.21	-0.24***	-3.87
ENROLSCH	-0.41*	-2.54	-0.09	-0.55
상수	-0.39**	-3.23	-0.80***	-3.99
Chi-Square	45.16		47.18	
df	7		7	
Log-likelihood	1141.18		908.48	
N	861		861	
종속변수 평균 (지속적인 전일제 근로자)	.55		.24	

\*P<.05 \*\*P<.01 \*\*\*P<.001

〈표 2〉 지속적인 전일제 근로자인 기혼남녀의 임금률에 대한 회귀분석

독립변수	기혼남성				기혼여성			
	교정(X)		교정		교정(X)		교정	
	B	t	B	t	B	t	B	t
HSGRAD	0.15	1.63	0.28	1.84	0.08	0.66	0.13	0.72
VOCED	0.37	3.17	0.49***	4.56	0.29	1.31	0.35	1.07
COLGRAD	0.42***	6.23	0.53***	4.15	0.53**	3.03	0.57**	3.18
PROF	0.63***	5.18	0.61***	5.61	0.49**	3.07	0.57**	2.92
WEEKWORK	0.001*	2.59	0.001*	2.40	0.00	0.51	0.001	1.18
JOBTRN	-0.03	-0.61	-0.01	-0.35	0.07	1.35	0.09	1.58
OVERTIME	-0.15***	-4.54	-0.17***	-4.79	-0.06	-0.94	-0.07	-1.04
AGE	0.02	1.78	0.02	1.99	0.04**	2.65	0.04**	2.84
COLBAR	0.16***	4.39	0.19***	4.62	0.06	1.13	0.002	0.04
URBAN	0.13***	4.04	0.15***	4.11	-0.01	-0.24	-0.02	-0.30
SOUTH	-0.07	-1.15	-0.04	-1.10	-0.06	-1.37	-0.07	-1.43
PUBLIC	-0.12*	-2.27	-0.16**	-2.75	0.03	0.36	0.01	0.16
CHILD6	-0.01	-0.41	0.07	0.78	0.02	0.57	-0.02	-0.12
DEPENDNT	-0.02	-0.31	-0.01	-0.36	0.04	-0.46	-0.08	-0.22
LAMBDA			-0.73	-0.83			-0.20	-0.08
상수	1.23*	2.34	1.82*	2.57	-0.35	-0.67	0.74	0.37
F	13.23		12.89		8.15		7.26	
R2(Adj)	0.258		0.28		0.304		0.32	
N			470		204			

\*P&lt;.05 \*\*P&lt;.01 \*\*\*P&lt;.001

루어져 있다. 첫째, 취업과 관련된 위험률을 추정하는 것이 필요하다. 이것은 지속적인 전일제 근로자 여부(가능성)를 추정하는 것이기 때문에 등식 (2)에 대한 회귀등식이 아닌 Probit 등식으로 절단되지 않은 표본을 추정한다. 이 등식에서 종속변수는 지속적인 전일제 근로자여부가 되며 독립변수는 교육, 6세이하 자녀 유무, 부양가족수, 학교재학여부 등을 포함한다. Probit 분석결과는 〈표 1〉에 나타나 있다. 기혼남성의 경우(55%)가 기혼여성(24%) 비해 지속적인 전일제 근로자가 되는 비율이 2배이상 높았다. 또한 기혼여성에 비해 더 많은 변수들이 기혼남성의 지속적인 전일제 근무자가 될 가능성을 높이거나 낮추는데 유의한 영향을 주고 있다.

일단 Probit 등식이 추정되면 그다음은  $\lambda$ 를 산출하기 위해 추정된 Probit계수를 이용하는 것이다. 이

Lambda는 절단된 표본의 회귀분석등식(등식 8)에 하나의 독립변수로 포함되어 다음과 같아진다.

$$\begin{aligned} \text{LN(WAGE)} = & \text{HSGRAD} + \text{VOCED} + \text{COLGRAD} \\ & + \text{PROF} + \text{WEEKWORK} + \text{JOBTRN} \\ & + \text{OVERTIME} + \text{AGE} + \text{COLBAR} \\ & + \text{URBAN} + \text{SOUTH} + \text{PUBLIC} \\ & + \text{CHILD6} + \text{DEPENDNT} \\ & + \text{LAMBDA}(\lambda) \end{aligned} \quad (9)$$

두가지 등식 (8), (9)의 추정결과가 〈표 2〉에 나타나 있다. 우선 열핏보기에 결정계수(R<sup>2</sup>)가 향상되었고 회귀계수들 사이에 몇가지 차이가 있음을 알 수 있다. 기혼남성의 경우 교정된 등식에서 통계학적으로 유의한 변수가 하나 더 추가 되었다. 독립변

수 하나와 종속변수의 상관관계가 역의 관계를 보여서 교정되지 않은 등식에서는 6세이하의 자녀가 있을 때 임금이 낮아지는 반면 교정된 등식에서는 오히려 임금이 높아지는 효과가 있는 것으로 나타났다. 기혼여성의 경우에는 두 등식에서 통계학적으로 유의한 변수의 수는 동일하였지만 상수와 한 독립변수의 종속변수에 대한 영향이 역의 방향으로 나타나 있다. 또한 남녀 모두에 있어서 통계학적으로 유의한 변수들(특히 학력)의 회귀계수들이 교정한 경우와 하지않은 경우에 차이를 보이고 있다.

자동차 운전자의 안전벨트 착용빈도에 관한 연구(Zick, 1985)에서는 무작위로 추출한 표본에서 운전할 못하는 응답자(총 응답자의 5%)를 제외시킨 편의를 교정을 하였다. 교정된 회귀분석결과 9개의 회귀자 중에서 6개의 회귀자의 통계학적 유의도가 향상되었다. 얼핏보기에 단지 5%의 표본절단에 대한 편의 교정은 무시될 수 있다고 볼 수도 있다. 그러나 실제로는 선택등식(Probit 등식)과 관심을 둔 등식(회귀등식)에 있는 오차항 사이의 상관관계의 정도만큼 추정치는 달라진다. 또한 회귀등식에 포함된 Lambda가 통계학적으로 유의하게 나타나서 표본선택편의에 대한 존재를 나타내고 있다.

## V. 표본중도절단을 교정한 기타 연구들

경제학자들에 의해 주목을 받아온 Heckman의 2단계 교정방법은 취업자의 시장임금이 아닌 비취업자의 기회비용접근법에 의한 임금추정시 많이 이용되어 왔다. 특히 비취업주부에 대한 임금추정연구에 가장 많이 이용되었고(Ferber & Green, 1985; Wang, 1990; Chiang, 1993; Puang & Metzen, 1993), 남편의 임금추정이나(Bryant & Wang, 1990) 노인의 임금추정(Kolodinsky & Walsh, 1991)의 경우 실업자에 의한 선택편의를 교정하기 위해 Probit 분석을 하였다. 이는 자체선택이 임금률에 영향을 준다고 널리 인식되어 있기 때문인데 그 이유는 취업할 가능성이 높은 사람들에게는 전통적으로 이용되는 변수로는 완전히 파악되지 않는 어떤 공통적인 특징이 있으리라 생각되기 때문이다. 즉 객관적인 자체선택은 취업자와 비

취업자간의 특성상의 차이를 말하고 주관적인 선택은 고용주가 갖는 고용인의 취업성취도에 대한 예측에서 의 특징을 말한다(Fan, 1993). 여기서 더 발전하여 최근에는 임금격차에 대한 연구에서도 이 교정방법을 이용한 연구들이 발표되고 있다(Nah, 1991; Fan, 1993).

가정관리분야에서는 가사노동의 가치산정연구시(Zick & Bryant, 1983, 1990; Duncan, 1992) 주부들의 취업여부에 따라 표본을 추출하여 발생하는 편의를 제외시키고 취업주부와 전업주부에 대한 연구결과를 일반화시키기 위해 Probit 분석을 하였다. 즉 시장임금이 가사노동임금보다 큰 취업여성에게서만 시장임금이 관찰될 수 있으므로 자체선택편의가 존재하며, 주부의 임금은 취업여부와 조직적으로 상관되어 있기 때문이다.

소비자분야에서는 주부의 취업여부에 따른 소비자출이나 소비행태연구(DeWeese & Norton, 1991; Douthitt & Fedyk, 1990; Kolodinsky & Labrecque, 1993) 등에서 임금추정시 선택편의의 교정을 하였다. 그외에 임금과는 관련없이 소비자의 불평행동연구(Kolodinsky, 1992)나 불충분한 정보를 가진 표본제외시 나타날 편의 가능성(O'Brien & Pritchard, 1989; Fast, Vosburgh, & Frisbee, 1989; Zick, 1985), 그리고 소비자의 선택에 따른 편의교정(Blaylock & Blisard, 1993; White-Means, 1992) Heckman의 방법이 이용되어 왔다.

## VI. Heckman의 교정방법에서 고려할 문제

현재까지 표본선택편의를 교정하는데 Heckman의 방법이 가장 널리 이용되고 있으나 몇가지 점에서 주의할 필요가 있다. 우선 회귀분석결과에 대한 의미가 자체선택변수에 의해 다른 일반 회귀분석결과와는 차이가 있다. 즉 회귀분석에 이용된 독립변수들이 선택등식에 포함된 변수들의 일부라면 추정된 회귀계수는 각 독립변수의 한계효과를 나타내는 것이 아니라 각 독립변수에 대한 조건부 한계효과를 나타낸다(Fan, 1993).

둘째로 Probit등식과 회귀등식에 포함된 변수중 적

어도 하나는 달라야 공선성(collinearity) 문제를 피할 수 있다(Maddala, 1983). 실제로는 자체선택과 관련된 모든 가능한 변수를 Probit 모델에 포함시키고 단지 의미있는 변수들만을 회귀모델에 포함시킨다. 그렇게 함으로써 추정된 자체선택 변수가 모델표기 오류에 의해 영향을 받지 않도록 할 수 있다.

세째로 회귀분석결과 Lambda가 통계학적으로 유의하다는 것은 표본선택편의가 존재함을 나타내며 또한 표본으로 선택된 집단과 선택되지 못한 집단간의 뚜렷한 특성의 차이가 있음을 반영하는 것이다. 실제로 본 연구에 인용된 연구들 중 약 60% 정도에서 Lambda가 유의하게 나타났는데 선택에 관련된 요인들이 얼마만큼 probit 등식에 포함되느냐와도 관계가 있다.

마지막으로 Heckman의 교정방법이 기회비용집근법에 의한 비취업자의 임금추정에 유용하지 못하다고 의문을 제시한 연구들(Zick & Bryant, 1983; Ferber & Green, 1985; Puang & Metzen, 1993)에 의하면 전업주부의 경우 낮은 기대소득때문이 아니라 높은 기회비용때문에 취업을 하지 않는 경우가 많아서 가사노동에 대한 최저가의 범위를 제시하는 추정계수보다는 연구대상자에게 직접 물어보는 것이 더 현실적이라고 설명하고 있다. 따라서 이부분에 대한 연구가 계속 진행되어야 하겠고 현재까지의 연구방법에 의한 추정된 계수와 직접 얻어진 정보를 비교분석하여 더 나은 연구방법을 모색해보아야 하겠다.

## VII. 결 론

경제학자들은 여러가지 다양한 상황에 의한 간접 표본중도절단을 교정하기위해 Heckman의 방식을 사용해왔다. 실제로 기혼여성에 대한 임금추정에서는 이 교정방법이 거의 자동적으로 이용되고 있다. 심리학이나 사회학, 정치학 등의 분야에서도 소수의 연구자들이 이러한 방법을 시도하고 있다. 아직까지 그 외의 다른 분야에서는 표본중도절단에 의해 일어나는 통계학적인 문제와 그 해결책에 대한 인식이 매우 미약한 편이다.

소비자연구나 가정관리에 관한 연구들이 종종 비확률부분표본들(nonrandom subsamples)을 이용하기 때문에 많은 분석들이 중도절단문제에 의해 방해를 받고 연구결과를 일반화시키는데 제한을 받는다. 비확률표본을 이용하는 연구자들은 표본중도절단이 적절한 것인지의 문제를 고려하여 선택과정이 모델화될 수 있다면 그리고 선택등식과 원래의 등식 사이의 오차항간의 상관관계가 0 이 아니라고 믿을만한 이유가 있다면 선택편의에 의한 교정으로 Heckman 방식을 이용해보는 것이 좋겠다. 표본선택편의가 일어난 경우 교정함으로서 실증연구에 대한 내적, 외적타당성을 현저히 향상시킬 수 있고 무엇보다도 왜곡된 연구결과가 이어지는 우려가 없어지기 때문이다. 또한 최근에는 이 교정절차를 간단히 실행할 수 있는 LIMDEP(Greene, 1988), TSP, RATS 등의 소프트웨어가 개발되어 있어서 이용자가 늘고 있다.

## 【참 고 문 헌】

- 1) Becker, G.S., "Human capital, effort, and the sexual division of labor", *Journal of Labor Economics* 3 (1), 1985:S33-S58.
- 2) Berk, R., "An introduction to sample selection bias in sociological data", *American Sociological Review* 48, 1983:386-398.
- 3) Blaylock, J.R. & Blisard, W.N., "Women and the demand for alcohol: Estimating participation and consumption" *JCA* 27(2), 1993:319-334.
- 4) Bryant, W.K. & Wang, Y., "Time together, time apart: An analysis of wives' solitary time and shared time with spouses" *Lifestyles: Family and Economic Issues* 11(1), 1990:89-119.
- 5) Chiang, F.F., "Household food-away-from-home and frozen food consumption and prices of wives' and husbands' time: A disaggregated cross-sectional analysis" *ACCI Proceedings* 39, 1993:66-73.
- 6) DeWeese, G. & Norton, M.J.T., "Impact of married women's employment on individual house-



- hold member expenditures for clothing" JCA 25(2) : 235-257.
- 7) Douthitt, R.A. & Fedyk, J.M., "Family composition, parental time, and market goods : Life cycle trade-offs" JCA 24(1), 1990:110-133.
  - 8) Duncan K.A., " The value of time in household work : estimates from the NLS data" ACCI Proceedings 38, 1992:163-170.
  - 9) Fan, X.J., "Self selection and wage gaps between genders and races" ACCI Proceedings 39, 1993: 7-65.
  - 10) Fast, J., Vosburgh, R.E., & Frisbee, W.R., "The effects of consumer education on consumer search" JCA 23(1), 1989:65-90.
  - 11) Ferber, M. & Green, C., "Homemakers' imputed wages : Results of the Heckman technique compared with women's own estimates", JHR 20, 1985:90-99.
  - 12) Greene, W., Limdep, Statistical package for estimating limited dependent variable equations, 1986.
  - 13) Heckman, J., " Sample selection bias as a specification error", Econometrica 45, 1979: 153-161.
  - 14) Kolodinsky, J., "Complaints, redress, & subsequent purchase of medical services by dissatisfied consumers", ACCI Proceedings 38, 1992:221-229.
  - 15) Kolodinsky, J. & Labrecque, J., "Shopping time, grocery expenditures, and coupon savings : Insights into a time/money tradeoff" ACCI Proceedings 39, 1993:162-171.
  - 16) Kolodinsky, J., & Walsh, R.W. " Estimating price effects in the absence of price data : a comparison of Elderly and Non-elderly households" ACCI Proceedings 37, 1991:134-143.
  - 17) Maddala, G.S., Limited dependent & qualitative variables in econometrics, Cambridge University Press, 1983.
  - 18) Nah, M., The male-female wage gap : a test of Becker's hypothesis, Unpublished dissertation, Ohio State University, 1991.
  - 19) O'Brien T.V. & Pritchard, M.R., "Development of a causal model of household expenditures for food consumed outside the home" ACCI Proceedings 35, 1989:216-220.
  - 20) Puang, T.C. & Metzen, E.J., " Reservation wages : An empirical test of alternative measures" ACCI Proceedings 39, 1993:135-141.
  - 21) Tuma, N.B., "Nonparametric & partially parametric approaches to event-history analysis in sociological methodology", Edited by S. Leinhardt, 1982:1-60.
  - 22) Wang, Y., "The shadow price of time : An empirical model incorporating involuntary unemployment" ACCI Proceedings 36, 1990:189-195.
  - 23) White-Means, S.I., "Allocation of labor to informal home health production : Health care for elderly, if time permits", JCA 26(1), 1992:69-89.
  - 24) Zick, C.D., "Sample selection bias : correcting for variables that aren't there", ACCI Proceedings, 1985:65- 71.
  - 25) Zick, C.D. & Bryant, W.K. "Shadow wage assessments of the value of home production : Patterns from the 1970's", Lifestyles : Family and Economic Issues 11(2), 1990:143-160.
  - 26) 한국통계학회, 통계용어사전, 자유아카데미, 1991.